

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants: Tomohiro Konoma : Art Unit:
Serial No.: To be Assigned : Examiner:
Filed: Herewith :
FOR: METHOD AND APPARATUS :
FOR SPEECH RECOGNITION :

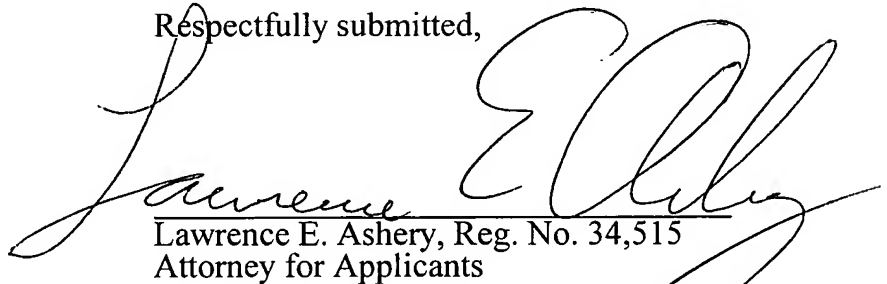
CLAIM TO RIGHT OF PRIORITY

Assistant Commissioner for Patents
Washington, D.C. 20231
S I R :

Pursuant to 35 U.S.C. 119, Applicants' claim to the benefit of filing of prior Japanese Patent Application No. 2001-014160, filed January 23, 2001, is hereby confirmed.

A certified copy of the above-referenced application is enclosed.

Respectfully submitted,



Lawrence E. Ashery, Reg. No. 34,515
Attorney for Applicants

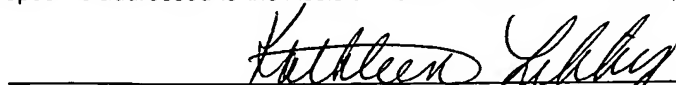
LEA/ebf
Enclosure: (1) certified priority document
Dated: January 17, 2002
Suite 301, One Westlakes, Berwyn
P.O. Box 980, Valley Forge, PA 19482
(610) 407-0700

The Assistant Commissioner for Patents is hereby authorized to charge payment to Deposit Account No. 18-0350 of any fees associated with this communication.

EXPRESS MAIL Mailing Label Number: EL 743541121 US

Date of Deposit: January 17, 2002

I hereby certify that this paper and fee are being deposited, under 37 C.F.R. § 1.10 and with sufficient postage, using the "Express Mail Post Office to Addressee" service of the United States Postal Service on the date indicated above and that the deposit is addressed to the Assistant Commissioner for Patents, Washington, D.C. 20231.



Kathleen Libby

日 本 国 特 許 庁
JAPAN PATENT OFFICE

Jc971 U.S. PTO
10/052145
01/17/02

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日

Date of Application:

2001年 1月23日

出 願 番 号

Application Number:

特願2001-014160

出 願 人

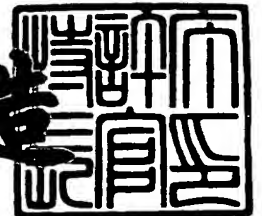
Applicant(s):

松下電器産業株式会社

2001年 9月19日

特 許 庁 長 官
Commissioner,
Japan Patent Office

及 川 耕 造



出証番号 出証特2001-3086323

【書類名】 特許願

【整理番号】 2931020125

【提出日】 平成13年 1月23日

【あて先】 特許庁長官殿

【国際特許分類】 G10L 15/22

【発明者】

【住所又は居所】 神奈川県川崎市多摩区東三田3丁目10番1号 松下技
研株式会社内

【氏名】 小沼 知浩

【発明者】

【住所又は居所】 神奈川県川崎市多摩区東三田3丁目10番1号 松下技
研株式会社内

【氏名】 井上 剛

【発明者】

【住所又は居所】 神奈川県川崎市多摩区東三田3丁目10番1号 松下技
研株式会社内

【氏名】 遠藤 充

【発明者】

【住所又は居所】 神奈川県川崎市多摩区東三田3丁目10番1号 松下技
研株式会社内

【氏名】 齋藤 夏樹

【発明者】

【住所又は居所】 神奈川県横浜市港北区綱島東4丁目3番1号 松下通
信工業株式会社内

【氏名】 石田 明

【発明者】

【住所又は居所】 神奈川県横浜市港北区綱島東4丁目3番1号 松下通
信工業株式会社内

【氏名】 木村 達也

【特許出願人】

【識別番号】 000005821

【氏名又は名称】 松下電器産業株式会社

【代理人】

【識別番号】 100097445

【弁理士】

【氏名又は名称】 岩橋 文雄

【選任した代理人】

【識別番号】 100103355

【弁理士】

【氏名又は名称】 坂口 智康

【選任した代理人】

【識別番号】 100109667

【弁理士】

【氏名又は名称】 内藤 浩樹

【手数料の表示】

【予納台帳番号】 011305

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9809938

【書類名】 明細書

【発明の名称】 音声認識方法及びその装置

【特許請求の範囲】

【請求項 1】 音素や音節などの音声単位の連結で表現され、単語間の接続ルールによって入力音声の長さに従い展開される仮説に対応する音響モデルと、入力音声とを照合し認識スコアを得、その認識スコアの高い少なくとも 1 個の仮説を残しながら入力音声に近い候補を認識結果として出力する音声認識方法において、入力された音声の特徴量をフレーム単位で抽出する特徴量抽出ステップと、現処理フレームが単語終端か単語内かを判別し、前記判定が単語終端ならば単語終端の仮説を記憶するか否かを判定し、単語内ならば単語内の仮説を記憶するか否かの判定をするステップと、前記記憶すると判定された仮説を記憶するステップと、前記記憶された仮説を仮説が単語内ならば単語を表現する音声単位を伸長し、単語終端ならば単語間の接続ルールにより次に続く単語を結合し、仮説を展開する制御を行う制御ステップと、前記入力音声から抽出されたフレーム特徴量と前記展開された仮説の音響モデルのフレーム特徴量との類似度を計算するステップと、前記類似度と前フレームまでの仮説の認識スコアとから、認識スコアの演算を行う演算ステップと、前記仮説と前記認識スコアを新しい仮説として登録するステップと、前記フレーム単位の処理を入力音声の終端まで続け認識スコアの高い少なくとも 1 個の仮説を入力音声に近い認識結果として出力するステップとを有することを特徴とする音声認識方法。

【請求項 2】 記憶するか否かの判定は、単語内仮説を記憶するか否かの判定の基準を類似度のスコアとし、単語終端仮説を記憶するか否かの判定の基準を仮説の個数としたことを特徴とする請求項 1 記載の音声認識方法。

【請求項 3】 音素や音節などの音声単位の連結で表現され、単語間の接続ルールによって入力音声の長さに従い展開される仮説に対応する音響モデルと、入力音声とを照合し認識スコアを得、その認識スコアの高い少なくとも 1 個の仮説を残しながら入力音声に近い候補を認識結果として出力する音声認識装置において、入力された音声の特徴量をフレーム単位で抽出する特徴量抽出手段と、現処理フレームが単語終端か単語内かを判別し、前記判定が単語終端ならば単語終端

の仮説を記憶するか否かを判定し、単語内ならば単語内の仮説を記憶するか否かの判定をする単語終端判定手段と、前記記憶すると判定された仮説を記憶する仮説記憶手段と、前記記憶された仮説を仮説が単語内ならば単語を表現する音声単位を伸長し、単語終端ならば単語間の接続ルールにより次に続く単語を結合し、仮説を展開する制御を行うサーチ制御手段と、前記入力音声から抽出されたフレーム特徴量と前記展開された仮説の音響モデルのフレーム特徴量との類似度を計算する類似度計算手段と、前記類似度と前フレームまでの仮説の認識スコアとから、認識スコアの演算を行うサーチ演算手段と、前記仮説と前記認識スコアを新しい仮説として登録する単語仮説登録手段と、前記フレーム単位の処理を入力音声の終端まで続け認識スコアの高い少なくとも1個の仮説を入力音声に近い認識結果として出力する認識結果出力手段とを有することを特徴とする音声認識装置。

【請求項4】 単語終端判定手段は、単語内仮説を記憶するか否かの判定の基準を類似度のスコアとし、単語終端仮説を記憶するか否かの判定の基準を仮説の個数としたことを特徴とする請求項3記載の音声認識装置。

【請求項5】 音素や音節などの音声単位の連結で表現され、単語間の接続ルールによって入力音声の長さに従い展開される仮説に対応する音響モデルと、入力音声とを照合し認識スコアを得、その認識スコアの高い少なくとも1個の仮説を残しながら入力音声に近い候補を認識結果として出力する音声認識するためにコンピュータを、入力された音声の特徴量をフレーム単位で抽出する特徴量抽出手段と、現処理フレームが単語終端か単語内かを判別し、前記判定が単語終端ならば単語終端の仮説を記憶するか否かを判定し、単語内ならば単語内の仮説を記憶するか否かの判定をする単語終端判定手段と、前記記憶すると判定された仮説を記憶する仮説記憶手段と、前記記憶された仮説を仮説が単語内ならば単語を表現する音声単位を伸長し、単語終端ならば単語間の接続ルールにより次に続く単語を結合し、仮説を展開する制御を行うサーチ制御手段と、前記入力音声から抽出されたフレーム特徴量と前記展開された仮説の音響モデルのフレーム特徴量との類似度を計算する類似度計算手段と、前記類似度と前フレームまでの仮説の認識スコアとから、認識スコアの演算を行うサーチ演算手段と、前記仮説と前記認

識スコアを新しい仮説として登録する単語仮説登録手段と、前記フレーム単位の処理を入力音声の終端まで続け認識スコアの高い少なくとも1個の仮説を入力音声に近い認識結果として出力する認識結果出力手段、として機能させるための音声認識プログラム。

【請求項6】 音素や音節などの音声単位の連結で表現され、単語間の接続ルールによって入力音声の長さに従い展開される仮説に対応する音響モデルと、入力音声とを照合し認識スコアを得、その認識スコアの高い少なくとも1個の仮説を残しながら入力音声に近い候補を認識結果として出力する音声認識するためにコンピュータを、入力された音声の特徴量をフレーム単位で抽出する特徴量抽出手段と、現処理フレームが単語終端か単語内かを判別し、前記判定が単語終端ならば単語終端の仮説を記憶するか否かを判定し、単語内ならば単語内の仮説を記憶するか否かの判定をする単語終端判定手段と、前記記憶すると判定された仮説を記憶する仮説記憶手段と、前記記憶された仮説を仮説が単語内ならば単語を表現する音声単位を伸長し、単語終端ならば単語間の接続ルールにより次に続く単語を結合し、仮説を展開する制御を行うサーチ制御手段と、前記入力音声から抽出されたフレーム特徴量と前記展開された仮説の音響モデルのフレーム特徴量との類似度を計算する類似度計算手段と、前記類似度と前フレームまでの仮説の認識スコアとから、認識スコアの演算を行うサーチ演算手段と、前記仮説と前記認識スコアを新しい仮説として登録する単語仮説登録手段と、前記フレーム単位の処理を入力音声の終端まで続け認識スコアの高い少なくとも1個の仮説を入力音声に近い認識結果として出力する認識結果出力手段、として機能させるための音声認識プログラムを記録したコンピュータ読み取り可能な記憶媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、一般の産業用、家庭用等の電気機器に搭載される音声認識技術に関し、特に音声認識の認識率向上を図る音声認識方法及び装置に関するものである。

【0002】

【従来の技術】

音声認識システムの一例として、「Hermann Ney: Data Driven Search Organization for Continuous Speech Recognition (IEEE TRANSACTIONS ON SIGNAL PROCESSING Vol.40 No.2 p272 1992)」が参照される。

【0003】

この音声認識システムは、図5の処理フローに示すステップを入力音声のフレームに同期して処理を行い、入力音声の終端まで実行することによって、入力音声に近い仮説を認識結果として得るものである。このような方法を用いるサーチをフレーム同期ビームサーチと呼ぶ。

【0004】

ステップ1：i番目のフレームの仮説はone-pass search アルゴリズムを用いて、i+1番目のフレームに展開する。仮説が単語内ならば単語を表現する音声単位を、単語終端ならば単語間接続ルールにより次に続く単語を結合しその最初の音声単位を伸長する。i番目のフレームの仮説は消去され、i+1番目の仮説だけが記憶される。

【0005】

ステップ2：i+1番目のフレームに展開された仮説のうち、最も累計スコアの高い仮説を基準として、そのスコアより一定の閾値以内のスコアを持つ仮説のみを記憶する。それ以外の仮説は消去する。これを候補の絞り込みといい、絞り込みにより、仮説の数が指数関数的に増加し計算不可能となることを避けている。

【0006】

ステップ3：処理すべきフレームiに1を加える。

【0007】

なお、フレーム同期ビームサーチにおいて高い精度で仮説の評価を行う音声認識方法として例えば特開平8-6588号公報には、フレーム同期ビームサーチにおいて時刻tの仮説のスコアを、時刻に対する正規化を行うために、全仮説に共通の尤度関数を差し引くことにより行っており、正規化されたスコアの最大値と各仮説の正規化されたスコアにより一定の閾値以内の仮説を記憶するものである。

【 0 0 0 8 】

【発明が解決しようとする課題】

しかしながら、上述した従来の音声認識システムにおいては、単語内の仮説も単語終端の仮説も、最も累計スコアの高い仮説を基準として、そのスコアより一定の閾値以内のスコアを持つ仮説のみを記憶するため、単語終端で後続単語候補が多数接続可能となり、仮説数の大幅な増大をまねき、効率的な絞り込みが行われず計算が困難な状態に陥るという課題があった。

【 0 0 0 9 】

したがって、本発明は、上記課題に鑑みてなされたものであって、その目的はフレーム同期ビームサーチにおいて、単語内では似た発音のスコアの高い仮説は単語終端まで、仮説数によらず精密に計算を行うことで認識精度を保ち、単語終端では、後続単語が接続することによる仮説数の増大を避けるために、仮説数による絞り込みを行い、認識精度を保ちつつ計算量を効果的に削減可能な音声認識方法及びその装置を提供することにある。

【 0 0 1 0 】

【課題を解決するための手段】

この課題を解決するために本発明は、音素や音節などの音声単位の連結で表現され、単語間の接続ルールによって入力音声の長さに従い展開される仮説に対応する音響モデルと、入力音声とを照合し認識スコアを得、その認識スコアの高い少なくとも1個の仮説を残しながら入力音声に近い候補を認識結果として出力する音声認識装置において、入力された音声の特徴量をフレーム単位で抽出する特徴量抽出手段と、現処理フレームが単語終端か単語内かを判別し、前記判定が単語終端ならば単語終端の仮説を記憶するか否かを判定し、単語内ならば単語内の仮説を記憶するか否かの判定をする単語終端判定手段と、前記記憶すると判定された仮説を記憶する仮説記憶手段と、前記記憶された仮説を仮説が単語内ならば単語を表現する音声単位を伸長し、単語終端ならば単語間の接続ルールにより次に続く単語を結合し、仮説を展開する制御を行うサーチ制御手段と、前記入力音声から抽出されたフレーム特徴量と前記展開された仮説の音響モデルのフレーム

特徴量との類似度を計算する類似度計算手段と、前記類似度と前フレームまでの仮説の認識スコアとから、認識スコアの演算を行うサーチ演算手段と、前記仮説と前記認識スコアを新しい仮説として登録する単語仮説登録手段と、前記フレーム単位の処理を入力音声の終端まで続け認識スコアの高い少なくとも1個の仮説を入力音声に近い認識結果として出力する認識結果出力手段とを備える。

【0011】

本発明は、仮説を記憶するか否かの判定を、現処理フレームが単語終端か単語内かを判別し、前記判定が単語終端ならば単語終端の仮説を記憶するか否かを判定し、単語内ならば単語内の仮説を記憶するか否かの判定をするようにしたものである。

【0012】

また、その判定基準を、単語内仮説を記憶するか否かの判定の基準を類似度のスコアとし、単語終端仮説を記憶するか否かの判定の基準を仮説の個数としたものである。

【0013】

【発明の実施の形態】

本発明の請求項1に記載の発明は、音素や音節などの音声単位の連結で表現され、単語間の接続ルールによって入力音声の長さに従い展開される仮説に対応する音響モデルと、入力音声とを照合し認識スコアを得、その認識スコアの高い少なくとも1個の仮説を残しながら入力音声に近い候補を認識結果として出力する音声認識方法において、入力された音声の特徴量をフレーム単位で抽出する特徴量抽出ステップと、現処理フレームが単語終端か単語内かを判別し、前記判定が単語終端ならば単語終端の仮説を記憶するか否かを判定し、単語内ならば単語内の仮説を記憶するか否かの判定をするステップと、前記記憶すると判定された仮説を記憶するステップと、前記記憶された仮説を仮説が単語内ならば単語を表現する音声単位を伸長し、単語終端ならば単語間の接続ルールにより次に続く単語を結合し、仮説を展開する制御を行う制御ステップと、前記入力音声から抽出されたフレーム特徴量と前記展開された仮説の音響モデルのフレーム特徴量との類似度を計算するステップと、前記類似度と前フレームまでの仮説の認識スコアと

から、認識スコアの演算を行う演算ステップと、前記仮説と前記認識スコアを新しい仮説として登録するステップと、前記フレーム単位の処理を入力音声の終端まで続け認識スコアの高い少なくとも1個の仮説を入力音声に近い認識結果として出力するステップとを有することにより、認識精度を保ちつつ計算量を効果的に削減できるという作用を有する。

【0014】

請求項2に記載の発明は、請求項1記載の音声認識方法において、記憶するか否かの判定は、単語内仮説を記憶するか否かの判定の基準を類似度のスコアとし、単語終端仮説を記憶するか否かの判定の基準を仮説の個数としたことにより、単語内では似た発音のスコアの高い仮説は単語終端まで、仮説数によらず精密に計算を行うことで認識精度を保ち、単語終端では、後続単語が接続することによる仮説数の増大を避けるために、仮説数による絞り込みを行い、認識精度を保ちつつ計算量を効果的に削減であるという作用を有する。

【0015】

請求項3に記載の発明は、音素や音節などの音声単位の連結で表現され、単語間の接続ルールによって入力音声の長さに従い展開される仮説に対応する音響モデルと、入力音声とを照合し認識スコアを得、その認識スコアの高い少なくとも1個の仮説を残しながら入力音声に近い候補を認識結果として出力する音声認識装置において、入力された音声の特徴量をフレーム単位で抽出する特徴量抽出手段と、現処理フレームが単語終端か単語内かを判別し、前記判定が単語終端ならば単語終端の仮説を記憶するか否かを判定し、単語内ならば単語内の仮説を記憶するか否かの判定をする単語終端判定手段と、前記記憶すると判定された仮説を記憶する仮説記憶手段と、前記記憶された仮説を仮説が単語内ならば単語を表現する音声単位を伸長し、単語終端ならば単語間の接続ルールにより次に続く単語を結合し、仮説を展開する制御を行うサーチ制御手段と、前記入力音声から抽出されたフレーム特徴量と前記展開された仮説の音響モデルのフレーム特徴量との類似度を計算する類似度計算手段と、前記類似度と前フレームまでの仮説の認識スコアとから、認識スコアの演算を行うサーチ演算手段と、前記仮説と前記認識スコアを新しい仮説として登録する単語仮説登録手段と、前記フレーム単位の処

理を入力音声の終端まで続け認識スコアの高い少なくとも1個の仮説を入力音声に近い認識結果として出力する認識結果出力手段とを有することにより、認識精度を保ちつつ計算量を効果的に削減できるという作用を有する。

【0016】

請求項4に記載の発明は、請求項3記載の音声認識装置において、単語終端判定手段は、単語内仮説を記憶するか否かの判定の基準を類似度のスコアとし、単語終端仮説を記憶するか否かの判定の基準を仮説の個数としたことにより、単語内では似た発音のスコアの高い仮説は単語終端まで、仮説数によらず精密に計算を行うことで認識精度を保ち、単語終端では、後続単語が接続することによる仮説数の増大を避けるために、仮説数による絞り込みを行い、認識精度を保ちつつ計算量を効果的に削減であるという作用を有する。

【0017】

請求項5に記載の発明は、音素や音節などの音声単位の連結で表現され、単語間の接続ルールによって入力音声の長さに従い展開される仮説に対応する音響モデルと、入力音声とを照合し認識スコアを得、その認識スコアの高い少なくとも1個の仮説を残しながら入力音声に近い候補を認識結果として出力する音声認識するためにコンピュータを、入力された音声の特徴量をフレーム単位で抽出する特徴量抽出手段と、現処理フレームが単語終端か単語内かを判別し、前記判定が単語終端ならば単語終端の仮説を記憶するか否かを判定し、単語内ならば単語内の仮説を記憶するか否かの判定をする単語終端判定手段と、前記記憶すると判定された仮説を記憶する仮説記憶手段と、前記記憶された仮説を仮説が単語内ならば単語を表現する音声単位を伸長し、単語終端ならば単語間の接続ルールにより次に続く単語を結合し、仮説を展開する制御を行うサーチ制御手段と、前記入力音声から抽出されたフレーム特徴量と前記展開された仮説の音響モデルのフレーム特徴量との類似度を計算する類似度計算手段と、前記類似度と前フレームまでの仮説の認識スコアとから、認識スコアの演算を行うサーチ演算手段と、前記仮説と前記認識スコアを新しい仮説として登録する単語仮説登録手段と、前記フレーム単位の処理を入力音声の終端まで続け認識スコアの高い少なくとも1個の仮説を入力音声に近い認識結果として出力する認識結果出力手段、として機能させ

るための音声認識プログラムをコンピュータで実行することにより、認識精度を保ちつつ計算量を効果的に削減できるという作用を有する。

【 0 0 1 8 】

請求項 6 に記載の発明は、音素や音節などの音声単位の連結で表現され、単語間の接続ルールによって入力音声の長さに従い展開される仮説に対応する音響モデルと、入力音声とを照合し認識スコアを得、その認識スコアの高い少なくとも 1 個の仮説を残しながら入力音声に近い候補を認識結果として出力する音声認識するためにコンピュータを、入力された音声の特徴量をフレーム単位で抽出する特徴量抽出手段と、現処理フレームが単語終端か単語内かを判別し、前記判定が単語終端ならば単語終端の仮説を記憶するか否かを判定し、単語内ならば単語内の仮説を記憶するか否かの判定をする単語終端判定手段と、前記記憶すると判定された仮説を記憶する仮説記憶手段と、前記記憶された仮説を仮説が単語内ならば単語を表現する音声単位を伸長し、単語終端ならば単語間の接続ルールにより次に続く単語を結合し、仮説を展開する制御を行うサーチ制御手段と、前記入力音声から抽出されたフレーム特徴量と前記展開された仮説の音響モデルのフレーム特徴量との類似度を計算する類似度計算手段と、前記類似度と前フレームまでの仮説の認識スコアとから、認識スコアの演算を行うサーチ演算手段と、前記仮説と前記認識スコアを新しい仮説として登録する単語仮説登録手段と、前記フレーム単位の処理を入力音声の終端まで続け認識スコアの高い少なくとも 1 個の仮説を入力音声に近い認識結果として出力する認識結果出力手段、として機能させるための音声認識プログラムを記録したコンピュータ読み取り可能な記憶媒体から音声認識プログラムをコンピュータにインストールして実行することにより、認識精度を保ちつつ計算量を効果的に削減できるという作用を有する。

【 0 0 1 9 】

以下、本発明の実施の形態について、図面を用いて説明する。

【 0 0 2 0 】

(実施の形態 1)

図 1 は、本発明の実施の形態 1 における音声認識装置の構成図の一例を示す図である。

【 0 0 2 1 】

図 1 において、1 0 1 は入力されたアナログ信号の音声からデジタル信号に変換する A/D 変換装置、1 0 2 はデータやプログラムを記憶する主記憶装置、1 0 3 はプログラムに従ってデータを処理する情報処理装置、1 0 4 は認識対象の単語を表現する音素や音節などの音声単位の音響的特徴をモデル化した複数フレームから構成される音響モデル、1 0 5 は認識対象の単語間の接続ルールを記述した言語モデル、1 0 6 はデータやプログラムを格納している外部記憶装置であり、1 6 1 は外部記憶装置 1 0 6 上にあり A/D 変換装置 1 0 1 によってアナログ音声からデジタル信号に変換された入力音声から音声の特徴量を抽出する特徴量抽出プログラム、1 6 2 は外部記憶装置 1 0 6 上にあり仮説が単語内か単語終端かを判定するプログラム、1 6 3 は外部記憶装置 1 0 6 上にあり 1 6 4 は認識スコアにより仮説を記憶するか否かを判定する単語内仮説記憶判定プログラムと、1 6 5 は仮説の個数により仮説を記憶するか否かを判定する単語終端仮説記憶判定プログラムからなる仮説記憶判定プログラム、1 6 6 は外部記憶装置 1 0 6 上にあり仮説記憶判定プログラムで記憶すると判定された仮説を記憶するプログラム、1 6 7 は外部記憶装置 1 0 6 上にあり仮説を仮説が単語内ならば単語を表現する音声単位を伸長し、単語終端ならば言語モデル 1 0 5 に記述された単語間の接続ルールにより次に続く単語を結合し、仮説を展開するフレーム同期ビームサーチの仮説の展開制御を行うサーチ制御プログラム、1 6 8 は外部記憶装置 1 0 6 上にあり特徴量抽出プログラム 1 6 1 から出力された入力音声のフレーム単位の特徴量と音響モデル 1 0 4 との類似度を計算する類似度計算プログラム、1 6 9 は外部記憶装置 1 0 6 上にあり類似度計算プログラム 1 6 8 により計算された類似度と前フレームまでの仮説の認識スコアとから認識スコアを計算するフレーム同期ビームサーチ演算プログラム、1 7 0 は外部記憶装置 1 0 6 上にあり仮説と計算された認識スコアを登録する仮説登録プログラム、1 7 1 は外部記憶装置 1 0 6 上にありフレーム同期ビームサーチを入力音声の終端まで続け認識スコアの高い出力可能な仮説を認識結果として出力する認識結果出力プログラム、1 0 7 は認識結果出力プログラム 1 7 1 の認識結果出力プログラムからの結果を出力する出力装置から構成されている。

【 0 0 2 2 】

図 2 は、音声認識装置の処理手順を表すフローチャートであり、このフローチャートを用いて、音声認識装置の処理手順を説明する。図中 S は各処理ステップを表し、各処理ステップは図 1 の特徴量抽出プログラム 1 6 1 ～認識結果出力プログラム 1 7 1 に対応するものである。

【 0 0 2 3 】

まず、入力されたアナログ信号の音声デジタル信号に A/D 変換する (S 2 1)。単語と認識スコアからなる仮説の初期集合を記憶する (S 2 2)。図 3 に示すように単語と認識スコアからなる仮説の初期集合を記憶する (S 2 2)。図 3 では、アアコン、アイシャドウ、アクセサリ、アクセス、アサガオ、イカスミという 6 つの単語 3 0 1 と認識スコア 3 0 2 の初期値を記憶した例である。

【 0 0 2 4 】

デジタル信号に変換されたの入力音声から特徴量をフレーム単位に抽出する (S 2 3)。仮説が現在単語内の音声単位を処理しているのか単語終端の音声単位を処理しているのかの判定を行う (S 2 4)。仮説の現処理フレームが単語終端ではなく単語内だった場合、最も認識スコアの高い単語内の仮説を基準として、その認識スコアより一定の閾値以内の認識スコアを持つ単語内の仮説のみに絞り込む (S 2 5)。仮説が単語終端である場合、認識スコアの高い順に仮説を選択し、仮説の個数により仮説を絞り込む (S 2 6)。絞り込まれた仮説が単語内ならば単語を表現する音声単位を伸長し、単語終端ならば単語間の接続ルールにより次に続く単語を結合し、新しい仮説として展開する (S 2 7)。

【 0 0 2 5 】

処理ステップ S 2 4 から S 2 7 までの処理進行について図 4 を用いて説明する。図 4 では、単語内の仮説の判定基準である最大認識スコアからの一定の閾値以内を 3、単語終端の仮説の判定基準である認識スコア上位 2 位までとして、○内の数値は求められた認識スコアとして説明する。

【 0 0 2 6 】

図 4 で S 2 2 によって登録された 6 つの単語に対してフレーム単位の処理を行い、時刻 t では認識スコアの値により単語仮説イカスミの認識スコアが閾値以下

になってしまうので、イカスミ以外の5つの単語が登録され処理が続けられる。このときの、最大認識スコアが12であり、その閾値は $12 - 3 = 9$ となる。時刻 $t+t1$ 時刻では、単語仮説アイシャドウの認識スコアが閾値以下になってしまうので、アアコン、アクセサリ、アクセス、アサガオが記憶され処理が続けられる。時刻 $t+t3$ では、アアコン、アクセス、アサガオの3つの単語が終端になり、アクセス、アサガオ、アアコンの順位になりアアコンは上位2位からまれ、アクセス、アサガオが記憶され処理が続けられるというフレーム同期ビームサーチ制御を行う。

【0027】

展開した仮説に対して、入力音声の特徴量と音響モデル104から類似度を計算し、さらに前フレームまでの仮説の認識スコアとから新しい認識スコアを計算するというフレーム同期ビームサーチ演算を行う(S28)。ここでは、認識スコアは、類似度と前フレームまでの仮説の認識スコアを加算して求めている。展開された仮説を新たな認識スコアとともに記憶する(S29)。

【0028】

終了判定条件が満たされるまで、上記S23からS29までのフレーム単位の処理過程を繰り返す(S30)。

【0029】

終了判定条件が満たされた時に残っていた仮説の集合から認識スコアの高い出力可能な仮説を認識結果として出力する(S31)。

【0030】

【発明の効果】

したがって、本発明によれば、フレーム同期ビームサーチにおいて、単語内では似た発音のスコアの高い仮説は単語終端まで、仮説数によらず精密に計算を行うことで認識精度を保ち、単語終端では、後続単語が接続することによる仮説数の増大を避けるために、仮説数による絞り込みを行い、認識精度を保ちつつ計算量を効果的に削減可能な音声認識方法及びその装置を提供することができる。

【図面の簡単な説明】

【図1】

本発明の実施の形態 1 による音声認識装置を示すシステム構成図

【図 2】

本発明の実施の形態 1 による音声認識装置の処理手順を示すフローチャート

【図 3】

本発明の実施の形態 1 において最初に登録される単語と認識スコアの初期集合を説明する図

【図 4】

本発明の実施の形態 1 に置ける仮説判定の処理進行を示す図

【図 5】

従来技術による処理手順を示すフローチャート

【符号の説明】

1 0 1 A/D変換装置

1 0 2 主記憶装置

1 0 3 情報処理装置

1 0 4 音響モデル

1 0 5 言語モデル

1 0 6 外部記憶装置

1 0 7 出力装置

1 6 1 特徴量抽出プログラム

1 6 2 単語内単語終端判定プログラム

1 6 3 仮説記憶判定プログラム

1 6 4 単語内仮説記憶判定プログラム

1 6 5 単語終端仮説記憶判定プログラム

1 6 6 仮説記憶プログラム

1 6 7 フレーム同期ビームサーチ制御プログラム

1 6 8 類似度計算プログラム

1 6 9 フレーム同期ビームサーチ演算プログラム

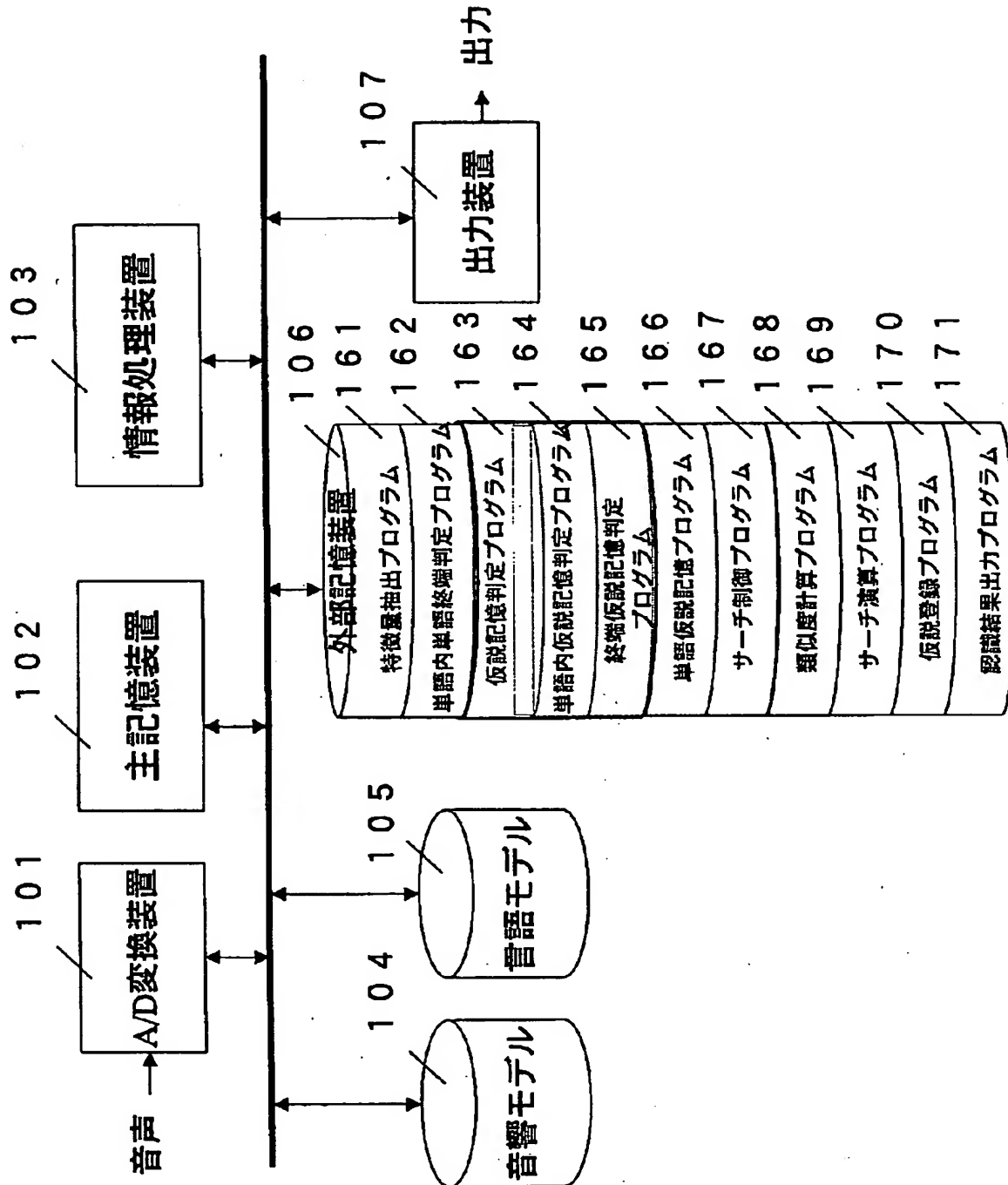
1 7 0 仮説登録プログラム

1 7 1 認識結果出力プログラム

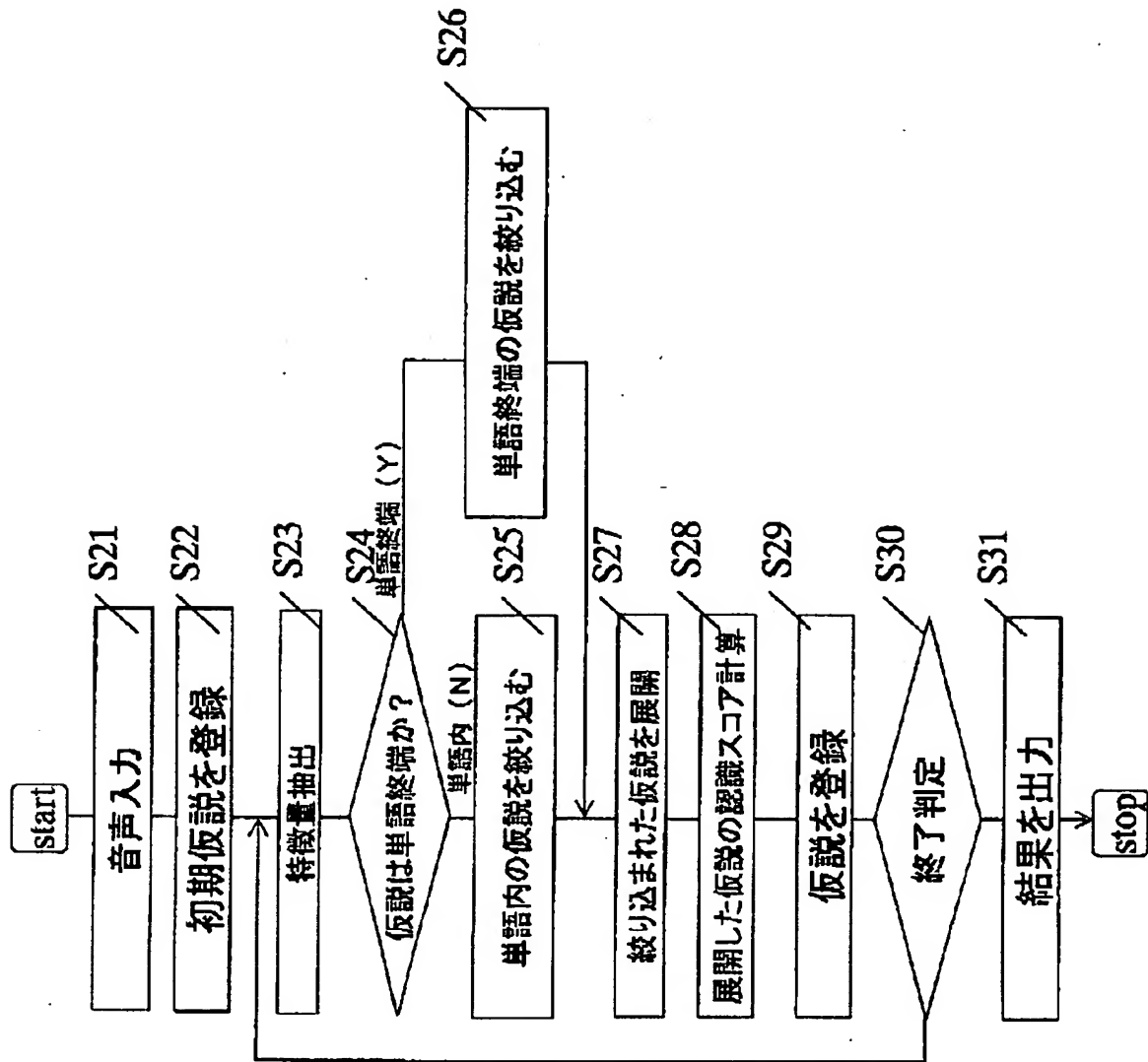
【書類名】

図面

【図 1】



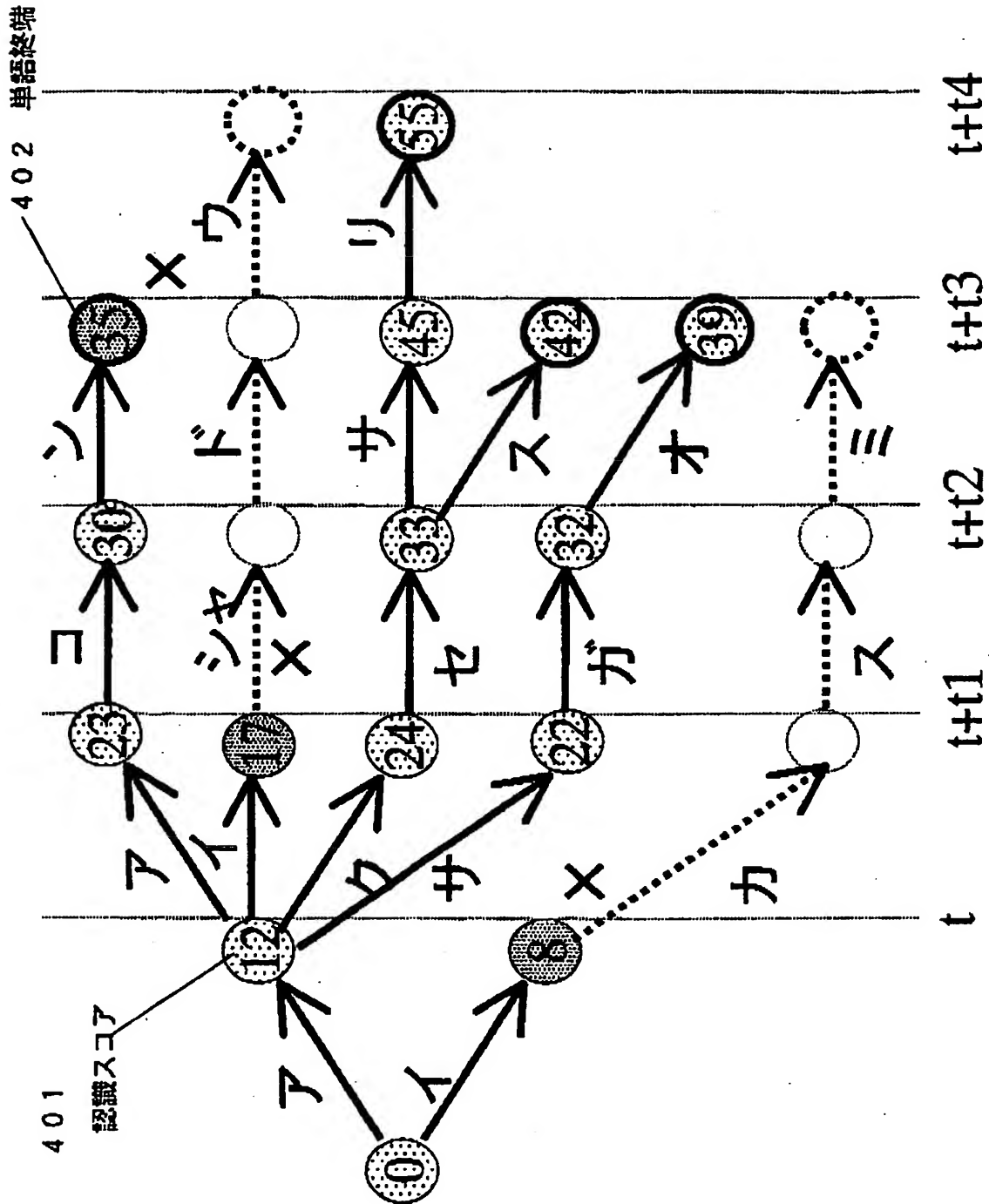
【図 2】



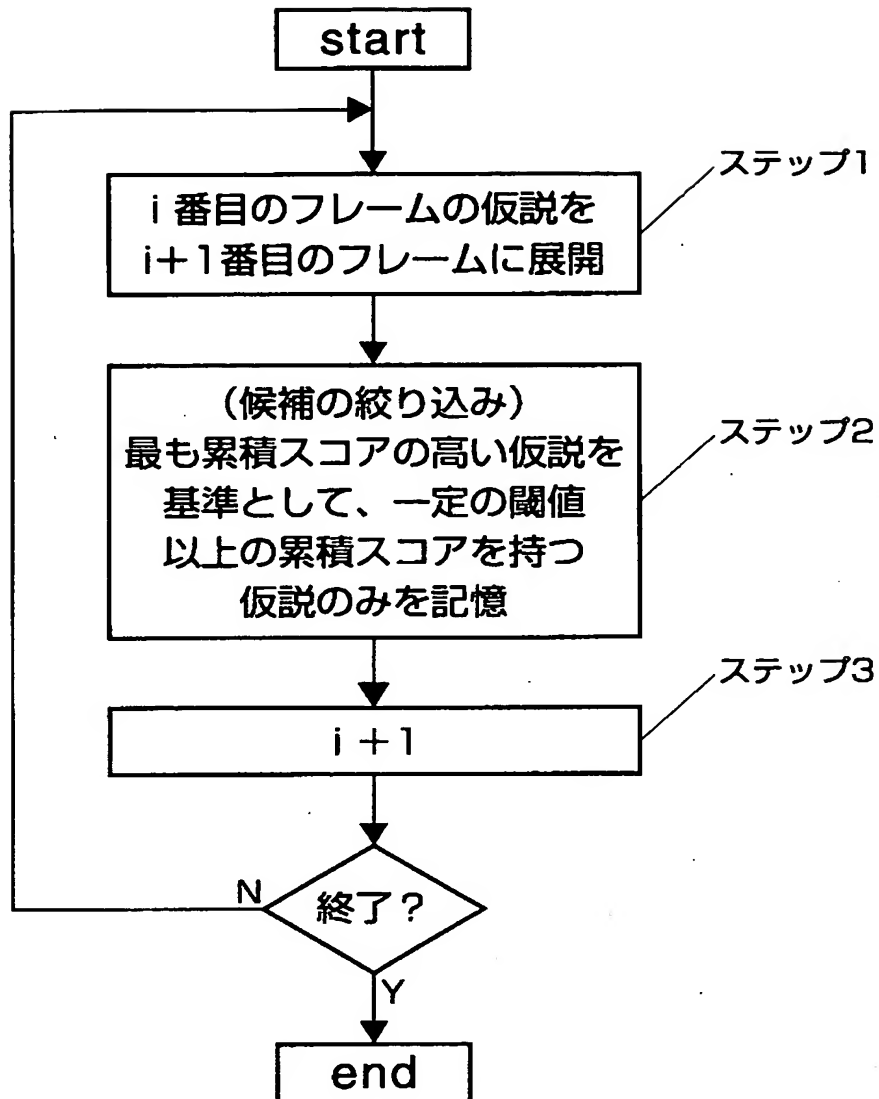
【図 3】

301 単語		・		302 認識スコア
		・		
		・		
		アアコン	0	
		アイシャドウ	0	
		アクセサリ	0	
		アクセス	0	
	アサガオ	0		
	イカスミ	0		
		・		
		・		
		・		

【図 4】



【図 5】



【書類名】 要約書

【要約】

【課題】 本発明は、認識精度を保ちつつ計算量を効果的に削減可能な音声認識方法及びその装置を提供することを目的とする。

【解決手段】 音素や音節などの音声単位の連結で表現され、単語間の接続ルールによって入力音声の長さに従い展開される仮説に対応する音響モデルと、入力音声とを照合し認識スコアを得、その認識スコアの高い少なくとも1個の仮説を残しながら入力音声に近い候補を認識結果として出力する音声認識装置において、単語内では似た発音のスコアの高い仮説は単語終端まで、仮説数によらず精密に計算を行うことで認識精度を保ち、単語終端では、後続単語が接続することによる仮説数の増大を避けるために、仮説数による絞り込みを行い、認識精度を保ちつつ計算量を効果的に削減可能な音声認識方法及びその装置を提供することができる。

【選択図】 図1

出 願 人 履 歴 情 報

識別番号 [000005821]

1. 変更年月日	1990年 8月28日
[変更理由]	新規登録
住 所	大阪府門真市大字門真1006番地
氏 名	松下電器産業株式会社